

## CLAIMS

What is claimed is:

1. A method of performing speaker verification, the method comprising:
  - a) obtaining a plurality of frames of compressed audio formants representing the speaker uttering a predetermined pass phrase, each frame including:
    - i) energy and pitch data characterizing the residue of the speaker uttering the predetermined pass phrase; and
    - ii) formant coefficients characterizing the resonance of the speaker uttering the predetermined pass phrase; and
  - b) verifying the identity of the speaker by matching at least one of:
    - i) energy data
    - ii) pitch data; and
    - iii) formant coefficientsin the frames to at least one of energy, pitch, and formant coefficients of a plurality of sample frames stored in memory.
2. The method of performing speaker verification of claim 1, wherein the step of obtaining frames of compressed audio formants includes receiving compressed audio formants from a remote Internet telephony device.
3. The method of performing speaker verification of claim 2, wherein the step of obtaining compressed audio formants from the remote Internet telephony device comprises receiving audio input of the speaker uttering the pass phrase from a microphone, digitizing the audio input, converting the digitized audio input to a sequence of frames of compressed audio formants, further compressing the sequence of frames of compressed audio formants to generate compressed audio data packets, and sending the compressed audio data packets from the remote Internet telephony device.

9  
1 4. The method of performing speaker verification of claim 3, wherein the step of  
2 verifying the identity of the speaker further includes normalizing the sequence of  
3 frames of compressed audio formants with the plurality of sample frames stored in  
4 memory utilizing at least one of the loudness, pitch, and formant data.  
5

1 5. A method of determining whether a speaker is a registered speaker, the  
2 method comprising:

3 a) obtaining compressed audio formants representing the speaker uttering  
4 a predetermined pass phrase, the compressed audio formants including:

5 i) energy and pitch data characterizing the residue of the speaker  
6 uttering the predetermined pass phrase;

7 ii) formant coefficients characterizing the resonance of the speaker  
8 uttering the predetermined pass phrase;

9 b) determining whether the speaker is the registered speaker by matching  
10 at least one of energy, pitch, and formant coefficients from the compressed audio  
11 formants to predetermined combinations of at least one of energy, pitch, and formant  
12 coefficients of sample compressed audio formants known to represent the registered  
13 speaker.  
14

1 6. The method of determining whether a speaker is a registered speaker of claim  
2 5, wherein the step of obtaining compressed audio formants includes obtaining the  
3 compressed audio formants from a remote location and sending the compressed  
4 audio formants from the remote location.  
5

1 7. The method of determining whether a speaker is a registered speaker of claim  
2 6, wherein the step of obtaining compressed audio formants at a remote location  
3 includes receiving audio input of the speaker uttering the pass phrase from a  
4 microphone, digitizing the audio input, and compressing the digitized audio input to  
5 generate compressed audio formants.

6

1 8. The method of determining whether a speaker is a registered speaker of claim  
2 7, wherein the compressed audio formants are a sequence of frames and each  
3 frame includes an energy value and a pitch value characterizing the residue of the  
4 speaker uttering the predetermined pass phrase and a plurality of formant  
5 coefficients characterizing the resonance of the speaker uttering the predetermined  
6 pass phrase and wherein the sample compressed audio formants are a sequence of  
7 frames and each frame includes an energy value and a pitch value characterizing the  
8 residue of the registered speaker uttering the predetermined pass phrase and a  
9 plurality of formant coefficients characterizing the resonance of the registered  
10 speaker uttering the predetermined pass phrase.

11

1 9. The method of determining whether a speaker is a registered speaker of claim  
2 8, wherein the step of determining whether the speaker is the registered speaker  
3 further includes normalizing the sequence of frames of compressed audio formants  
4 with the sequence of frames of sample compressed audio formants within the time  
5 domain.

6

1 10. A speaker verification server for verifying the identity of a remote speaker, the  
2 server comprising:

- 3 a) a network interface for receiving compressed audio formants via a  
4 packet switched network representing a remote speaker uttering a predetermined  
5 pass phrase as audio input to a remote telephony client;  
6 b) a database storing a plurality of compressed audio formant samples,  
7 each representing a registered speaker uttering a registered pass phrase as audio  
8 input; and  
9 c) a verification application operatively coupled to each of the network  
10 interface and the database for comparing the compressed audio formants  
11 representing the remote speaker to a compressed audio formant sample to  
12 determine whether the remote speaker is the registered speaker.

13

1 11. The speaker verification server of claim 10, wherein the compressed audio  
2 formants include energy and pitch data characterizing the residue of the speaker  
3 uttering the predetermined pass phrase and formant coefficients characterizing the  
4 resonance of the speaker uttering the predetermined pass phrase; and each  
5 compressed audio formant sample includes energy and pitch data characterizing the  
6 residue of the registered speaker uttering the registered pass phrase and formant  
7 coefficients characterizing the resonance of the registered speaker uttering the  
8 registered pass phrase.

9

1 12. The speaker verification server of claim 11, wherein the verification application  
2 determines whether the at least one of energy, pitch, and formant coefficients from  
3 the compressed audio formants is similar to the at least one of the energy, pitch, and  
4 formant coefficients of the sample.

5

1 13. The speaker verification server of claim 12, wherein the compressed audio  
2 formants are a sequence of frames and each frame includes an energy value, a pitch  
3 value, and a plurality of formant coefficients representing a portion of the utterance of  
4 the speaker and the compressed audio formant sample is a sample sequence of  
5 frames and each frame includes an energy value, a pitch value, and a plurality of  
6 formant coefficients representing a portion of the utterance of the registered speaker.

7

1 14. The speaker verification server of claim 13, wherein the verification application  
2 determines whether the sequence of frames is similar to the sample sequence of  
3 frames by comparing energy, pitch, and formant coefficients from each frame in the  
4 sequence of frames to energy, pitch, and formant coefficient from a corresponding  
5 frame in the sample sequence of frames.

6

1 15. The speaker verification server of claim 14, wherein the verification application  
2 further normalizes the sequence of frames with the sample sequence of frames  
3 within the time domain.  
4

1 16. A telephony server comprising:

2 a) a network interface for sending and receiving compressed audio  
3 formants to and from each of a plurality of telephony clients;

4 b) a telephony server application for maintaining a telephony  
5 session between an initiating telephony client and a terminating subscriber loop by:

6 i) receiving compressed audio formants from the initiating  
7 telephony client, decompressing the compressed audio formants to generate  
8 an audio signal, and sending the audio signal to the terminating subscriber  
9 loop; and

10 ii) receiving an audio signal from the terminating subscriber  
11 loop, compressing the audio signal to compressed audio formants, and  
12 sending the compressed audio formants to the telephony client;

13 c) a database storing a plurality of compressed audio formant  
14 samples, each representing one of a plurality of authorized users uttering a  
15 registered pass phrase; and

16 d) a verification application operatively coupled to each of the  
17 network interface and the database for comparing compressed audio formants  
18 received from the telephony client with at least one of the plurality of compressed  
19 audio formant samples to determine whether an operator of the telephony client is an  
20 authorized user.  
21

1 17. The telephony server of claim 16, wherein the compressed audio formants  
2 include energy and pitch data characterizing the residue of the speaker uttering the  
3 predetermined pass phrase and a plurality of formant coefficients characterizing the  
4 resonance of the speaker uttering the predetermined pass phrase; and each  
5 compressed audio formant sample includes energy and pitch data characterizing the

6 residue of the registered speaker uttering the registered pass phrase and a plurality  
7 of formant coefficients characterizing the resonance of the registered speaker  
8 uttering the registered pass phrase.  
9

1 18. The telephony server of claim 17, wherein the verification application  
2 determines whether at least one of energy, pitch, and formant coefficients from the  
3 compressed audio formants is similar to at least one of the energy, pitch, and  
4 formant coefficients of a compressed audio formant sample.  
5

1 19. The telephony server of claim 18, wherein the compressed audio formants are  
2 a sequence of frames and each frame includes an energy value, a pitch value, and a  
3 plurality of formant coefficients representing a portion of the utterance of the speaker  
4 and each compressed audio formant sample is a sample sequence of frames and  
5 each frame includes an energy value, a pitch value, and a plurality of formant  
6 coefficients representing a portion of the utterance of the registered speaker.  
7

1 20. The telephony server of claim 19, wherein the verification application  
2 determines whether the sequence of frames is similar to the sample sequence of  
3 frames by comparing at least one of energy, pitch, and formant coefficients in each  
4 frame to at least one of energy, pitch, and formant coefficients in a corresponding  
5 frame from the sample sequence of frames.  
6

1 21. The telephony server of claim 20, wherein the verification application further  
2 normalizes the sequence of frames with the sample sequence of frames within the  
3 time domain.  
4